*Indian Journal of Animal Sciences* **82** (4): 392–396, April 2012

# Genome wise identification of long terminal repeats in *Bubalus bubalis*

INDU KHATRI[1], UPASNA SHARMA[2], PRIYANKA BANERJEE[3], JYOTI JOSHI[4] and R K VIJH[5]

*National Bureau of Animal Genetic Resources, Karnal, Haryana 132 001 India*

Received: 23 May 2011; Accepted: 29 October 2011

ABSTRACT

Bovines are important livestock species contributing towards food (milk and meat), draft power and fuel. Buffalo is important for Indian subcontinent and far East countries of Asia and Brazil. Striking difference between the cattle and buffalo is attributed to distinct novelties between them at genomic level. Long terminal repeat (LTR) retrotransposons make up a large fraction of the typical mammalian genome. The abundance of LTR retro-transposons are believed to hold major significance for genome structure and function. The present study is an effort to identify the contribution of LTR to *Bubalus bubalis* genome, their abundance and the distribution of LTRs was examined in the buffalo genome version Bbu_2.0–alpha using web based algorithm. The buffalo chromosomes were screened against tRNA database of *Bos taurus*. A total of 26,338 LTRs were observed from the buffalo genome. Chromosome 1 exhibited the maximum number of LTRs *i.e.* 2232 and the minimum *i.e.* 324 LTRs were observed in chromosome 24. It was observed that more LTRs were present on sense strand than on antisense strand. 11 LTR regions were found to be coding for the Reverse Transcriptase (RT). Target site repeats (TSR) were AT rich. The density of the LTR along the length of chromosomes revealed that the LTR are present abundantly in the middle of autosomes than at the end except chromosome X. Being source of genetic variability and having evolutionary significance, an indepth analysis of LTRs may provide a clue for the significant difference between cattle and buffalo genomes.

**Key words**: Buffalo, Chromosomes, Long terminal repeat, Reverse transcriptase

The long terminal repeats (LTR) retrotransposons exist in the eukaryotic genomes (Ganko *et al.* 2001, Kaptionov and Jurka *et al.* 2003) and have been analysed in *Bos taurus* (Almeida *et al.* 2007). The LTRs have been reported to contribute to host gene structure, function and expression (Han and Boeke 2004, Dunn *et al.* 2005, 2006 and Ganko 2006). All the studies have shown that the transposable elements (TE) insertions in the genome have diverse evolutionary consequences. The transposable elements have several cis-regulatory components which include the promoters and enhancer sequences and thus capable of not only influencing their own activity but also expression of adjacent genes (Dunn *et al.* 2003, 2005, 2006). The TEs may also lead to evolution of coding regions and thus lead to genetic novelties (Ganko *et al.* 2003 and Britten 2004).

Next generation sequencing technologies for DNA are opening new vistas and are providing unprecedented chances to explore the function and evolutionary impact using large scale genetic information (Paterson *et al.* 2003, Zhang and Wessler 2004). The study of LTRs is important for the species which contribute to the human welfare by providing food (milk and meat), draft and fuel. The analysis of transposable elements was carried out after the sequence information of *Bos taurus* was publically made available (Almeida *et al.* 2007). The other bovine with whole genome sequence assembly is buffalo. The analysis of the buffalo genome includes reports for transposons *i.e.*, LINEs and SINEs (Tantia *et al.* 2011). In this paper, we carried out LTR analysis of the buffalo genome using complete sequence for all the 24 autosomes and X chromosome.

## MATERIALS AND METHODS

*Data collection*

The buffalo genome assembly was downloaded from http://210.212.93.84/cgi-bin/gb2/gbrowse/bovine. The Fasta files of 24 autosomes and X chromosome were scanned using LTR_Finder for finding the LTR regions incorporated into the buffalo genome. We utilized web server which provided an efficient means for the prediction of full length LTR retrotransposons available at http://tilife.fudan.edu.cn/ltr_finder/(Xu and Wang 2007). We divided each chromosome into smaller sequences of approximately ~40 Mb size owing to the maximum limit of 50 Mb of the webserver.

LTR_Finder applies rapid algorithms to construct reliable

Present address: [1,3,4]Senior Research Fellow ([1] indu2287@gmail.com), [2] Research Associate and [5] Principal Scientist.

LTRs and to predict accurate element boundaries through a multi-refinement process. The tRNA database of *Bos taurus* (Baylor release Btau_4.0 Oct, 2007) was taken for prediction of primer binding sites. All other parameters were kept as default. The distribution of LTR was analyzed chromosomes wise. The numbers of LTRs were analysed for both the DNA strands i.e. sense and antisense strand were recorded. Among the various LTR present in the genome the LTRs coding for Reverse Transcriptase (RT) were also identified. Practical Extraction and Reporting Language (PERL) was used to fetch out the information of score, length, tRNA type, Target site repeats (TSR) strands on which the LTR were present.

## RESULTS AND DISCUSSION

The score threshold for the LTR was kept at 6. The score measured the signals (TSR, TG.CA box, PBS, PPT, $IN_{(core)}$, $IN_{(c\text{-}term)}$, RT, RH). There are a total of 11 signals and thus the score varies from 0–11. The minimum score obtained was 6 and maximum 8. Most of the LTRs were observed with score 6 and 7 with no LTR coding for RT, Integrase and RNase H.

The buffalo chromosomes were scanned by LTR_Finder program for the presence of the number of LTR. The length of LTR was found to vary from 1218 to 26304 nucleotides. Various overlapping LTR were also observed in the output. A total of 26,338 LTR sequences were found to be incorporated in the buffalo genome. The maximum number of LTR i.e. 2232 were observed in chromosome 1 and least i.e. 324 in chromosome 24. The number of LTRs for each of the buffalo chromosome is depicted in Fig. 1. Chromosome 1 being the largest (205 Mbp) had the largest number of LTRs while the minimum number of were found in smallest chromosome 24 (42 Mbp). This may be attributed to the total genomic content of the chromosome.

The density of LTRs present along the length of chromosomes was plotted in order to find out the preferable integration of the LTR. The density scatter graph was plotted as number of LTR per 10 Mb of chromosomes sequence (Fig. 2). The LTR were sparsely present in the start and end of the chromosomes except Chromosome X where the LTR were equally distributed along the length of the chromosome and increased at the end of the chromosomes. LTRs were scarcely present at the beginning of the chromosomes and then suddenly increase before smoothening along the length of the chromosomes. The numbers of LTRs do not follow a specific trend however there is a steep decline in their number at the terminal end of each chromosome.

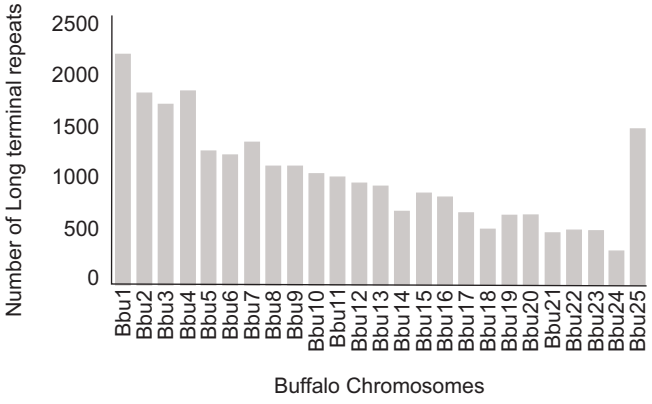



Fig. 1. Number of LTR present in buffalo (chromosome-wise *Bbu1-Bbu24* and *BbuX*).

The number of LTRs was compared in the sense and antisense strands for their presence. The number of LTRs were varying in both strands i.e. sense and antisense strands. However, the percentage distribution of LTR is above 50% in sense strand and below 50% in antisense strand. The percentage distribution of LTR in sense strand and antisense strand is shown using a bar diagram (Fig. 2).

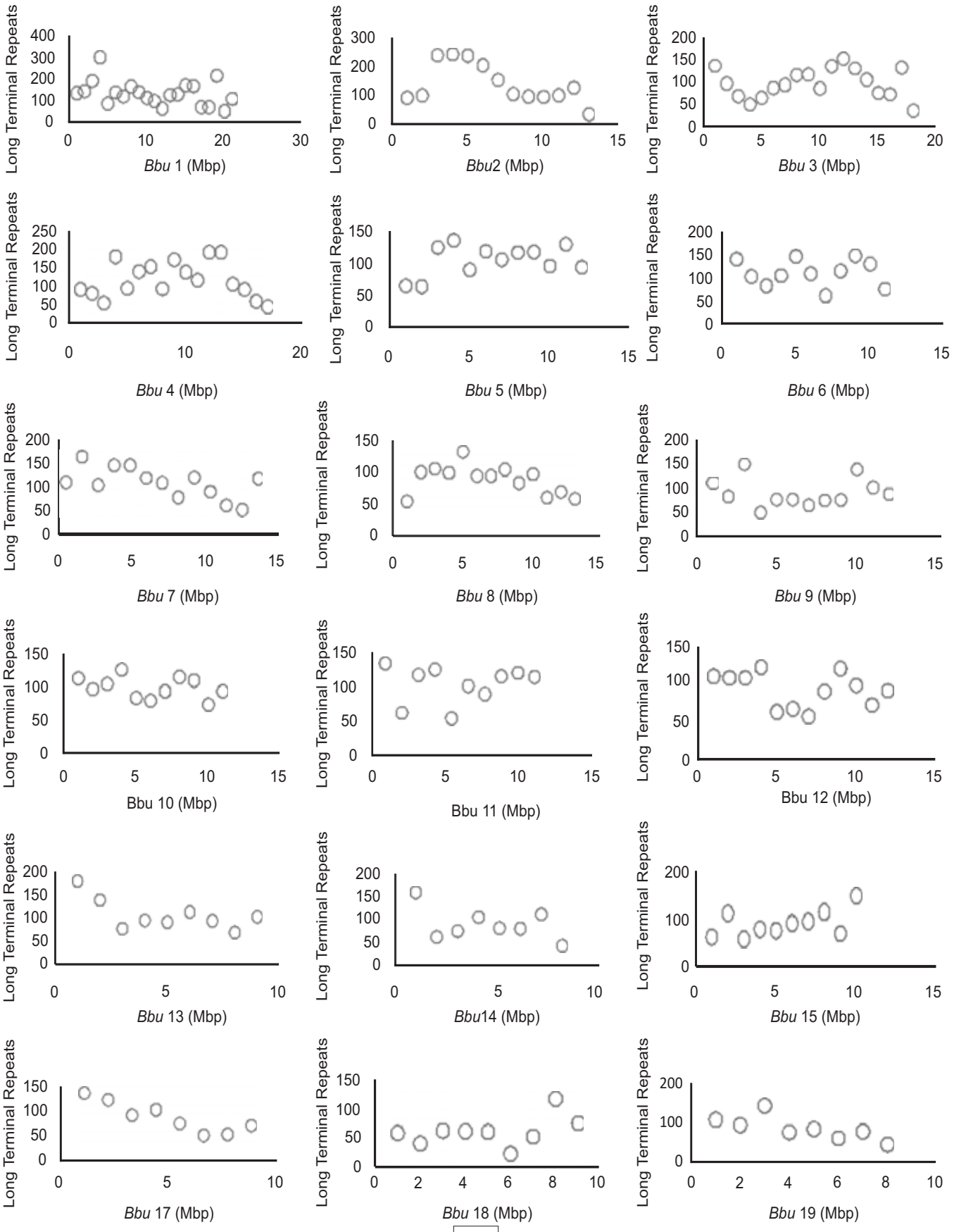
A comparison of presence of LTRs in the sense and antisense strands revealed that the number of LTRs varied in both strands i.e. sense and antisense strands. The percentage distribution of LTR in sense strand and antisense strand is shown as a bar diagram (Fig. 3). There were more LTRs in the sense strand compared to the antisense strands in all the buffalo chromosomes.

Target site repeats (TSR) was also identified in the LTR sequences. They were very less in number as compared to the number of LTRs (Fig. 4). The TSRs were observed to be biased towards more A/T than G/C *i.e.* the TSRs were AT rich.

The sequences with score 7.5 and 8 were analyzed for their status which revealed that these LTRs also code for RT but no LTR was found with coding region of Integrase and Rnase H. Only 11 LTR sequences were found to code for the RT. Among them 5 RTs were found in chromosomes *Bbu2*, *Bbu3*, *Bbu5*, *Bbu13*, *Bbu15*, and *6 RTs* were found in *Bbu6*, *Bbu18* and *BbuX* i.e. two in each chromosome. A typical output of the LTR finder is depicted in (Fig. 5). RT is an element important for the transposition of transposable elements in retroviruses as it transcribes single stranded RNA into double stranded DNA and this result in increase in their copy numbers.

RTs were equally distributed on sense and antisense strand viz., 6 on sense and 5 on antisense strands were observed. The RTs sequences were retrieved from the chromosome's fasta files and BLASTX was performed. The RT sequences were found to show similarity to the Reverse transcriptase like protein and hypothetical proteins of *Bos taurus* with 69% to 89% similarity. The chromosomes were scanned against the *Bos taurus* tRNA database. The LTRs present in the chromosomes preferred some tRNA types more than the others. CysGCA, GluTTC, GlyTCC, TrpCCA were the preferred tRNA followed by ValAAC, MetCAT, LysTTT, GluCTC,CysACA, AlaAGA.

The buffalo belong to the same family as that of cattle but the difference exist in the terms of production, fat and protein percentage and adaptability. The high fat and protein


Long Terminal Repeats
*Bbu* 1 (Mbp)
*Bbu*2 (Mbp)
*Bbu* 3 (Mbp)
*Bbu* 4 (Mbp)
*Bbu* 5 (Mbp)
*Bbu* 6 (Mbp)
*Bbu* 7 (Mbp)
*Bbu* 8 (Mbp)
*Bbu* 9 (Mbp)
Bbu 10 (Mbp)
Bbu 11 (Mbp)
Bbu 12 (Mbp)
*Bbu* 13 (Mbp)
*Bbu*14 (Mbp)
*Bbu* 15 (Mbp)
*Bbu* 17 (Mbp)
*Bbu* 18 (Mbp)
*Bbu* 19 (Mbp)

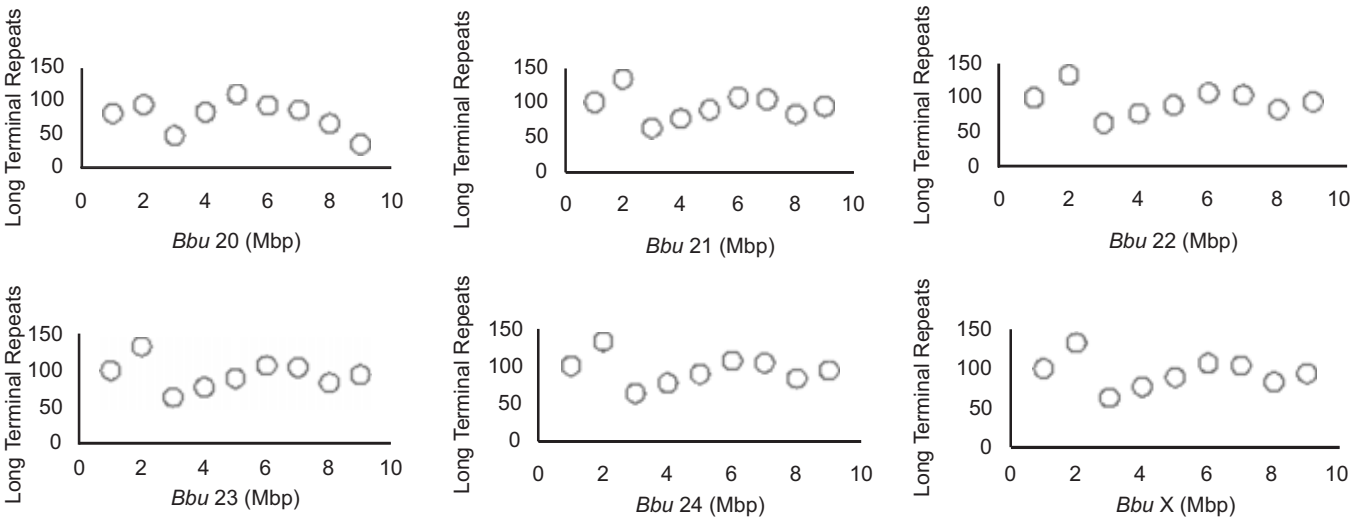



Fig. 2. Density of the LTRs along the length of the buffalo chromosomes.

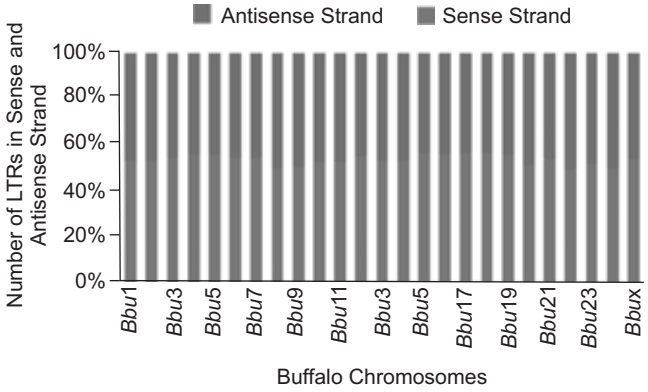



Fig. 3. Percentage distribution of LTR in sense and antisense strands of buffalo chromosomes.

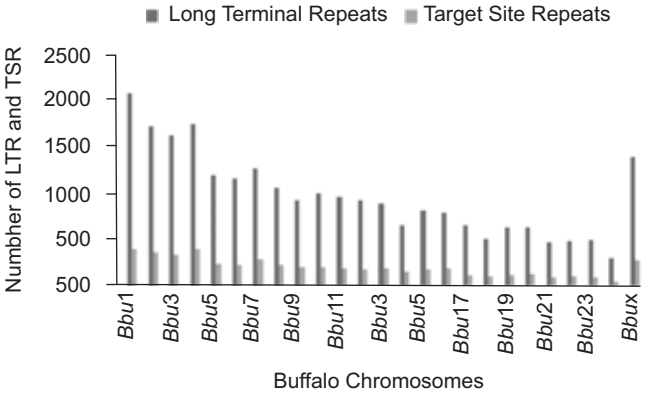



Fig. 4. Chromosome-wise distribution of number of LTRs and TSRs in buffalo.

percentage and several other important economic attributes of the buffalo may be the genetic novelties and results of evolutionary and functional biology. The transposable elements are the one of the major source of the genetic variability and also provide DNA template for the novel genes

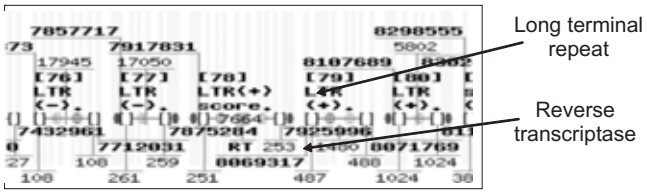



Fig. 5. Result of LTR_Finder showing the LTR and Reverse transcriptase (RT) of the region of buffalo chromosome.

or regulatory sequences. The TE-gene associations in cattle (Almeida *et al.* 2007) have shown that a large number of genes harbor insertions of transposable elements of different length and a small fraction of them are inserted into the coding sequences. The TE insertions may lead to different gene products by alternative splicing and thus change the function. Such a situation may also exist in buffalo. However, the present study was not restricted to genes but has taken the whole genome sequence. As the buffalo sequence assembly was an alignment and not a *de novo* assembly, similarities in results for cattle and buffalo are expected. Although TE fragments modify genomes and contribute to host gene expression, deeper biochemical and molecular studies are needed to discern the phenomenon.

## REFERENCES

Almeida L M, Silva I T, Silva W A, Castro J P, Riggs P K, Carareto C M and Amaral E J. 2007. The contribution of transposable elements to *Bos taurus* gene structure. *Gene* **390**: 180–89.

Britten R J. 2004. Coding sequence of functioning human genes derived entirely from mobile element sequences. *Proceedings of the National Academy of Sciences of the USA* **101**:16825–30.

Dunn C A, Medstrand P and Mager D L. 2003. An endogenous retroviral long terminal repeat is the dominant promoter for human beta1, 3–galactosyltransferase 5 in the colon.

*Proceedings of the National Academy of Sciences of the USA* **100**: 12841–46.

Dunn C A, van de Lagemaat L N, Baillie G J and Mager D L. 2005. Endogenous retrovirus long terminal repeats as ready-to-use mobile promoters: the case of primate beta3GAL-T5. *Gene* **364**: 2–12.

Dunn C A, Romanish M T, Gutierrez L E, van de Lagemaat L N and Mager D L. 2006. Transcription of two human genes from a bidirectional endogenous retrovirus promoter. *Gene* **366**: 335–42.

Ganko E W, Fielman K T and McDonald J F. 2001. Evolutionary history of *Cer* elements and their impact on the *C. elegans* genome. *Genome Research* **11**: 2066–74.

Ganko E W, Bhattacharjee V, Schliekelman P and McDonald J F. 2003. Evidence for the contribution of LTR retrotransposon to *C. elegans* gene evolution. *Molecular Biology and Evolution* **20**:1925–31.

Ganko E W, Greene C S, Lewis J A, Bhattacharjee V and McDonald J F. 2006. LTR retrotransposon-gene associations in Drosophila melanogaster. *Journal of Molecular Evolution* **62** (1): 111–20.

Han J S and Boeke J D. 2004. A highly active synthetic mammalian retrotransposon. *Nature* **429**(6989): 314–18.

Paterson A H, Bowers J E, Peterson D G, Estill J C and Chapman B A. 2003. Structure and evolution of cereal genomes. *Current Opinion in Genetics and Development* **13**: 644–50.

Tantia M S, Vijh R K, Bhasin V, Sikka P,Vijh P K, Kataria R S, Mishra B P, Pandy A K, Yadav S P, Sethi R K, Joshi B K, Gupta S C and Pathak K M L. 2011.Whole-genome sequence assembly of the water buffalo (*Bubalus bubalis*). *Indian Journal of Animal Sciences* **81**(5) 38–46.

Xu Z and Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acid Research* **35**: W265–68.

Zhang X and Wessler S R. 2004. Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. *Proceedings of the National Academy of Sciences of the USA* **101**(15): 5589–94.